CLAIMS:

1. A method for generating image data for rendering on a hardcopy document, comprising:

identifying a primary set of symbol data; the primary set of symbol data providing a first channel of human readable information to be rendered on the hardcopy document; and

computing a secondary set of encoding data from the primary set of symbol data; the secondary set of encoding data providing an assist channel of machine readable information to be rendered on the hardcopy document; wherein said computing further comprises:

partitioning the primary set of symbol data into a plurality of groups; the plurality of groups classifying symbol data according to how likely the symbol data will occur in the hardcopy document, and

developing the secondary set of encoding data by associating the symbol data with ones of the plurality of groups.

2. The method according to claim 1, further comprising assigning an identifier to each of the plurality of groups that partition the primary set of symbol data.

3. The method according to claim 1, wherein the plurality of groups further classify the symbol data according to how likely symbol data in the primary set of symbol data are to be confused during processing of a scanned representation of the primary set of symbol data.

4. The method according to claim 1, further comprising compressing the secondary set of encoding data.

5. The method according to claim 4, further comprising scrambling the compressed secondary set of encoding data.

6. The method according to claim 5, wherein said compressing is performed with arithmetic encoding.

7. The method according to claim 4, wherein said scrambling further comprises:

computing a hash h_i for each symbol in the primary set of symbol data with a state change function $H(h_{i-1}, c_i)$, where h_{i-1} is a previous occurrence of the state change function H for the symbol c_{i-1}; and

developing the secondary set of encoding data cd_i for each corresponding symbol c_i by scrambling each separation code d_i with a guard value g_i derived from the computed hash h_i.

8. The method according to claim 1, further comprising encoding the machine readable information with data glyphs.

9. The method according to claim 1, wherein the primary set of symbol data and the secondary set of encoding data are rendered by a printer on the hardcopy document.

10. The method according to claim 1, wherein said partitioning further comprises:

defining a graph with a node for each symbol in the primary set of symbol data;

assigning a value to each of the nodes that quantifies how likely the symbol data will occur in the hardcopy document;

interconnecting the nodes of the graph with arcs; and

assigning a value to each of the arcs that quantifies how likely symbol data in the primary set of symbol data are likely to be confused during processing of the scanned representation of the symbol data.

11. The method according to claim 10, wherein said partitioning further comprises iteratively compute a partition of the graph that maximizes a ratio of

the values of each of the arcs and the values of each of the nodes until achieving a desired benefit or entropy.

12. The method according to claim 1, further comprising:

recording a scanned representation of the hardcopy document that includes the primary set of symbol data and a secondary set of encoding data; and

decoding the scanned representation of the hardcopy document by identifying a shortest path of a product graph of the scanned representation of the primary set of symbol data and the secondary set of encoding data.

13. The method according to claim 12, wherein said decoding is performed using a shortest path computation.

14. The method according to claim 13, wherein the shortest path computation comprises a two-pass dynamic programming computation.

15. The method according to claim 14, wherein the shortest path computation further comprise defining a product graph.

16. The method according to claim 15, wherein the product graph is defined by:

identifying templates from a template library that matches the primary set of symbol data;

defining nodes that correspond to both position in the primary set of symbol data and the secondary set of encoding data; and

defining arcs that identify template matches of the primary set of symbol data that are consistent with the secondary set of encoding data.

17. An apparatus for generating image data for rendering on a hardcopy document, comprising:

means for identifying a primary set of symbol data; the primary set of symbol data providing a first channel of human readable information to be rendered on the hardcopy document; and

means for computing a secondary set of encoding data from the primary set of symbol data; the secondary set of encoding data providing an assist channel of machine readable information to be rendered on the hardcopy document; wherein said computing means further comprises:

5   means for partitioning the primary set of symbol data into a plurality of groups; the plurality of groups classifying symbol data according to how likely the symbol data will occur in the hardcopy document, and

means for developing the secondary set of encoding data by associating the symbol data with ones of the plurality of groups.

10   18. The apparatus according to claim 17, wherein the plurality of groups further classify the symbol data according to how likely symbol data in the primary set of symbol data are to be confused during processing of a scanned representation of the primary set of symbol data.

19. The apparatus according to claim 17, wherein said partitioning further

15   comprises:

means for defining a graph with a node for each symbol in the primary set of symbol data;

means for assigning a value to each of the nodes that quantifies how likely the symbol data will occur in the hardcopy document;

20   means for interconnecting the nodes of the graph with arcs; and

means for assigning a value to each of the arcs that quantifies how likely symbol data in the primary set of symbol data are likely to be confused during processing of the scanned representation of the symbol data.

20. The apparatus according to claim 17, further comprising:

25   means recording a scanned representation of the hardcopy document that includes the primary set of symbol data and a secondary set of encoding data; and

means decoding the scanned representation of the hardcopy document by identifying a shortest path of a product graph of the scanned representation of the primary set of symbol data and the secondary set of encoding data.